CH5350: Applied Time-Series Analysis

Estimation of Time-Domain Statistical Properties

Arun K. Tangirala

Department of Chemical Engineering, IIT Madras

Introductory Remarks

- In practice, measurements always contain a stochastic component. Whether a deterministic component is present or not depends on the application and the assumptions made by the user.
- The signal-to-noise ratio (SNR) is a measure of the proportions in which these two components are present with respect to the variance of the signal.
- It is important to know if a particular estimation method is appropriate for the signal of interest since some methods (e.g., periodogram) are best suited only for deterministic signals while some others only for random signals.

Estimation of mean

Two widely used estimators of mean are the **sample mean** and the **sample median**.

Sample mean:

$$\boxed{\bar{y} = \sum_{k=0}^{N-1} \alpha_k y[k]} \qquad \text{where } \alpha_k = \frac{1}{N}, \ \forall\, k \tag{1}$$

Sample mean: Properties

- It is the MoM and OLS estimator. It is also the MLE when $y[k] \sim \text{GWN}(\mu, \sigma^2)$.
- The estimator is *unbiased* when $y[k]$ is *stationary*.
- Under the *stationarity* assumption for $\{y[k]\}$,

$$\text{var}(\bar{y}) = \sigma_{\bar{y}}^2 = \frac{1}{N} \sum_{l=-(N-1)}^{N-1} \left(1 - \frac{|l|}{N}\right) \sigma_{yy}[l] \tag{2}$$

- Sample mean is a **consistent** estimator of the mean for stationary $y[k]$.
- It is the most efficient estimator when $\{y[k]\}$ is Gaussian white. For correlated processes, a WLS estimate or MLE with a knowledge of $\sigma_{yy}[l]$ should be used.

Distribution of sample mean

The CLT establishes the asymptotic distribution of $\bar{y}$ under fairly general conditions.

Theorem

If $\{y[k]\}$ has a linear stationary representation

$$y[k] = \mu_y + \sum_{n=-\infty}^{\infty} h[n]w[k-n] \quad \textit{under} \qquad w[k] \sim \textit{i.i.d.}(0, \sigma_w^2) \tag{3}$$

then,

$$\boxed{\bar{y} \sim \textit{As.}\mathcal{N}(\mu, \frac{\lambda}{N}) \qquad \textit{where} \qquad \lambda = \sum_{l=-\infty}^{\infty} \left(1 - \frac{|l|}{N}\right) \sigma_{yy}[l]} \tag{4}$$

Distribution of sample mean

- For example, when $y[k]$ is WN, (4) specializes to $\bar{y} \sim \text{As}\mathcal{N}(\mu, \sigma^2/N)$, a known result. From where the *large sample* $100(1-\alpha)\%$ C.I. can be constructed

$$\boxed{\bar{y} - z_{\alpha/2}\frac{\sigma}{\sqrt{N}} < \mu < \bar{y} + z_{\alpha/2}\frac{\sigma}{\sqrt{N}}} \tag{5}$$

  where $z_{\alpha/2}$ is the critical value satisfying Pr$(Z > z_{\alpha/2} = \alpha/2)$ and $Z \sim \mathcal{N}(0,1)$.
- When σ is unknown, the *sample* standard deviation may be used.
- In the **small sample case** $(N \leq 50)$, the standardized sample mean $t = (\bar{y} - \mu)/(\hat{\sigma}/\sqrt{N})$ is known to follow a Student's t-distribution only when $y[k] \sim$ GWN. The critical values in the C.I. are then replaced with those from t-distribution.

Sample Median

The sample median is the middle (or a pseudo-middle) value of the *ordered* observations,

$$\mathbf{y}_s = \text{sort}(\mathbf{y}); \quad y_s[0] \geq y_s[2] \geq \cdots \geq y_s[N-1]$$

$$\breve{y} \triangleq \text{Sample Median}(\mathbf{y}) = \sum_{i=0}^{N-1} \alpha_i y_s[i] \tag{6}$$

$$\text{where } \alpha_i = \begin{cases} 1, & i = \frac{N+1}{2} \text{ (odd } N); \\ \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix}, & i = \frac{N}{2}, \frac{N}{2}+1 \text{ (even } N) \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

- Thus, sample median and sample mean can both be viewed as weighted estimators

Properties of sample median

- It is the optimal estimate of mean in the 1-norm sense,

$$\breve{y} = \arg\left\{ \min_c \sum_{k=0}^{N-1} |y[k] - c| \right\}$$

  Statistically it minimizes $E(|Y - c|)$ (when this expectation exists).
- It is an **unbiased** estimator of the mean whenever the density $f(y)$ is symmetric. It is also an unbiased estimator of the median, but w/o any restrictive assumptions.
- The sample median is asymptotically normal

$$\breve{y} \sim \mathcal{N}(\mu, \frac{1}{4Nf^2(m)}) \quad \text{where } m \text{ is the true median and } f(.) \text{ is the p.d.f..} \tag{8}$$

Properties of sample median

- Relatively, median is a *less efficient estimator of* μ than the sample mean for **Gaussian distributions**. The asymptotic relative efficiency is about $2/\pi = 0.6366$.
  - However, the situation is reversed when the signal belongs to other distributions. For example, with a Laplace distribution, the median is fully efficient.
- The lower efficiency of the sample median is balanced by its **high robustness**. It has a maximum breakdown point (Huber and Ronchetti, 2009) of 0.5, making it ideally suitable for estimating mean from data corrupted by outliers.
- The sample median is a **strongly consistent estimator of the population median** whenever $\{y[k]\}$ is i.i.d.

Estimation of variance: Sample variance

Two versions of sample variance estimators are:

$$\hat{\sigma}^2_{N-1} = \frac{1}{N-1}\sum_{k=0}^{N-1}(y[k]-\bar{y})^2; \qquad \hat{\sigma}^2_{N} = \frac{1}{N}\sum_{k=0}^{N-1}(y[k]-\bar{y})^2 \tag{9}$$

Properties of sample variance

- The estimators $\hat{\sigma}^2_{N-1}$ and $\hat{\sigma}^2_{N}$ are the OLS and ML (as well as MoM) estimators of the true variance σ^2 respectively.
- The ML estimator is **biased**, while the OLS estimator is unbiased. Needless to add, both are asymptotically unbiased.

Properties of sample variance

- The variance of each of these estimators when $\{y[k]\}$ is WN is given by

$$\text{var}(\hat{\sigma}^2_{N-1}) = \frac{2\sigma^4}{(N-1)^2}; \qquad \text{var}(\hat{\sigma}^2_N) = \frac{2(N-1)\sigma^4}{N^2} \tag{10}$$

- Thus, both estimators are consistent, but the ML estimate has a lower standard error and is efficient. Asymptotically they are equally efficient.
- The ML estimator is used more in practice. It is also commensurate with the widely used version of ACVF estimator (to be shortly introduced).
- *Distribution*: When $\{y[k]\}$ is *Gaussian* white, the normalized estimate has a χ^2 distribution with $N-1$ degrees of freedom:

Properties of sample variance ...contd.

- Distribution: $\dfrac{N\hat{\sigma}^2_N}{\sigma^2} \sim \chi^2_{N-1}$

  A $100(1-\alpha)$ C.I. for σ^2 can thus be derived,

$$\boxed{\frac{N\hat{\sigma}^2}{\chi^2_{N-1,1-\alpha/2}} < \sigma^2 < \frac{N\hat{\sigma}^2}{\chi^2_{N-1,\alpha/2}}} \tag{11}$$

  Due to the nature of the χ^2 distribution, the **C.I. is not symmetric**

  - When the distribution of $\{y[k]\}$ is non-normal, the above results approximately hold good provided the deviation from Gaussianity is not too serious.

Other estimators of spread: Range

Three other estimators of **spread** (standard deviation) are the **range**, **mean absolute deviation** and the **median absolute deviation**.

1. **Range**: The range is simply defined as,

$$R = |\max \mathbf{y} - \min \mathbf{y}| \tag{12}$$

   - For a normal distribution, $R \approx 6$ when N is large. The *interquartile range* (IQR), is a fairly robust estimator of σ.
   - For a normal distribution, the estimate of standard deviation is given by

$$\boxed{\hat{\sigma} = \text{IQR}/1.34898} \tag{13}$$

   where the correction factor is applied to obtain an unbiased estimate of σ.

Other estimators of spread: μAD

2. **Mean absolute deviation** (μAD): The theoretical definition is

$$\delta_1 \triangleq \mu\text{AD} = E(|Y - \mu)| \tag{14}$$

   - The μAD is viewed as a more "natural" way of expressing the spread of a random variable than the standard deviation.
   - Closed-form expressions for δ_1 do not generally exist, but expressions for the ratio δ_1/σ do exist.

μAD ... contd.

- A commonly used estimator of μAD is

$$\hat{\delta}_{1,y} = \frac{1}{N}\sum_{k=0}^{N-1} |y[k] - \bar{y}| \quad (15)$$

- A correction factor is usually necessary to obtain an unbiased estimate of the regular scale parameter σ. For normally distributed data,

$$\boxed{\hat{\sigma}_1 = 1.253\hat{\delta}_1} \quad (16)$$

- The resulting estimator is more robust than the standard sample variance.

Other estimators of spread: MAD

3. **Median absolute deviation:** (MAD): The definition of MAD is same as μAD, with the mean replaced by median in theory as well as in estimation.

$$\hat{\delta}_{2,y} \triangleq \text{MAD}(\mathbf{y}) = \frac{1}{N}\sum_{k=0}^{N-1} |y[k] - \breve{y}| \quad (17)$$

As with μAD, a correction factor is usually necessary. Once again, for normally distributed data, we have that

$$\hat{\sigma}_2 = 1.4826\hat{\delta}_2 \quad (18)$$

Properties of MAD

- This estimate is more robust than the μAD.
- Further, $\hat{\sigma}_2 \xrightarrow{p} \sigma$ with an asymptotic distribution $\mathcal{N}(\mu, \sigma^2)$.
- The efficiency of MAD is low for data falling out of a Gaussian distribution, standing at 37%.
- An alternative measure by Rousseeuw and Croux, 1993 uses the first quartile of all interpoint distances to give 82% efficiency at normal distribution, which is more than a notch higher than that of MAD.

See Pham-Gia and Hung, 2001 for a nice exposition on μAD and MAD.

Estimation of correlation

Computation of correlation requires covariance estimates.

A standard way of estimating covariance between two signals $y[k]$ and $u[k]$ is through

$$\boxed{\hat{\sigma}_{yu} = \frac{1}{N}\sum_{k=0}^{N-1}(y[k]-\bar{y})(u[k]-\bar{u})} \tag{19}$$

The correlation estimate (a.k.a. *Pearson's correlation coefficient*) is then given by

$$\boxed{\hat{\rho}_{yu} = \frac{\hat{\sigma}_{yu}}{\hat{\sigma}_y\hat{\sigma}_u}} \tag{20}$$

Estimation of correlation ...contd.

The covariance and correlation matrices are constructed accordingly:

$$\hat{\Sigma} = \begin{bmatrix} \hat{\sigma}_y^2 & \hat{\sigma}_{yu} \\ \hat{\sigma}_{uy} & \hat{\sigma}_u^2 \end{bmatrix}; \qquad \Xi = \begin{bmatrix} 1 & \hat{\sigma}_{yu} \\ \hat{\rho}_{uy} & 1 \end{bmatrix} \tag{21}$$

Properties of sample correlation coefficient

- The sample correlation coefficient is an **asymptotically unbiased and consistent** estimator of the correlation ρ_{yu}.s
- **Distribution**: The distribution of $\hat{\rho}$ is a rather complicated problem to handle. When $y[k]$ and $u[k]$ have a joint Gaussian distribution, an exact expression involving $beta$ and hypergeometric functions is available (Pearson, 2011).

Properties of sample corr. coeff. ...contd.

- Given the complicated nature of the result, typically approximate expressions for the first four moments are computed. Among these we are interested in the first two:

$$E(\hat{\rho}_{yu}) \approx \rho - \frac{\rho(1-\rho^2)}{2(N+6)}; \qquad \text{var}(\hat{\rho}_{yu}) \approx \frac{(1-\rho^2)^2}{N+6} \tag{22}$$

Observe that the variance of the estimate is maximum when $\rho = 0$, *i.e.*, when the variables are actually uncorrelated.

Properties of sample corr. coeff. ...contd.

- Under the bivariate Gaussian assumption, when the true correlation between the signals is $\rho = 0$, the estimate is known to have a near normal distribution.

$$\hat{\rho}_{yu} \xrightarrow{d} \mathcal{N}(0, 1/N) \tag{23}$$

Thus, the 95% and 99% **significance levels** for correlation are $\pm 1.96/\sqrt{N}$ and $\pm 2.58/\sqrt{N}$ respectively.

Properties of sample corr. coeff. ...contd.

- When the true $\rho_{yu} \neq 0$, Fisher's transformation produces a transformed coefficient with approximately normal distribution, under the large sample approximation and bivariate Gaussian assumption.

$$F_{\hat{\rho}} = \frac{1}{2}\ln\left(\frac{1+\hat{\rho}}{1-\hat{\rho}}\right) \sim \mathcal{N}(\mu_F, \sigma_F^2) \tag{24}$$

$$\text{where } \mu_F = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right); \quad \sigma_F^2 = \frac{1}{N-3} \tag{25}$$

R commands

Listing 1: R commands for estimating mean, variance and covariance

```

# Computing mean
mean, median, mode

# Computing variance and standard deviation
var, sd, cov, cor, range, IR, mad(y,center=median(y))
```

Estimation of CCVF

The estimate of the CCVF $\sigma_{yu}[l]$ between two signals y and u is the cross-covariance estimate between $y[k]$ and $u[k-l]$.

$$\boxed{\hat{\sigma}_{yu}[l] = \frac{1}{N}\sum_{k=l}^{N-l-1}(y[k]-\bar{y})(u[k-l]-\bar{u}) \quad l \geq 0} \tag{26}$$

- For negative lags, use $\sigma_{yu}[l] = \sigma_{uy}[-l]$.
- The maximum lag up to which $\hat{\sigma}_{yu}[l]$ can be evaluated is $|l_{\text{max}}| = N-1$.
- Observe that the estimate in (26) simplifies to the ML estimate of variance in (9).

Estimation of CCF

An estimate of the CCF is obtained by simply normalizing the CCVF estimate,

$$\boxed{\hat{\rho}_{yu}[l] = \frac{\hat{\sigma}_{yu}[l]}{\sqrt{\hat{\sigma}_{yy}[0]\hat{\sigma}_{uu}[0]}}} \tag{27}$$

Properties

The estimator in (27) is **asymptotically unbiased** and consistent. This is expected a correlation estimate is underlying (27).

Properties of CCF estimator

In TSA, we are primarily interested in, (i) *whiteness of residuals (or a given series)* and (ii) *lack of correlation between residuals and (lagged) inputs*.

Thus, the following is required to be known:

1. **Variability (or standard error) in $\hat{\rho}_{yu}[l]$ when the true correlation is $\rho_{yu}[l] = 0, \ \forall l$.** This is useful in drawing the significance levels for CCF (ACF) estimates.
2. **Correlation between estimates at two different lags, again when the true correlation function is zero.** This becomes useful in interpreting the CCF (ACF) plots.

Properties of CCF estimator

A general result is given in Brockwell and Davis, 1991; Shumway and Stoffer, 2006. We shall only study the case when the true correlation is zero and present the salient points.

- The significance levels for correlation estimates **when one of the signals is hypothesized to be white** are given as below:

$$\hat{\rho}_{yu}[l] \sim \text{As.}\mathcal{N}(0, 1/N) \tag{28}$$

$$\Longrightarrow 100(1-\alpha)\% \text{ sig. level for } \hat{\rho}_{yu}[l] = \pm\frac{z_{\alpha/2}}{\sqrt{N}} \tag{29}$$

Properties of CCF estimator . . . contd.

- **If neither signal is known to be white, then one or both have to be pre-whitened to be able to use the results**.

Pre-whitening: Given any series, the process first involves fitting a sufficiently high-order time-series model (preferably AR because of the ease of estimation) to the given series. The residuals of the resulting model is the pre-whitened series.

Example: Testing for zero cross-correlation

Zero cross-correlation test

Consider two signals that are theoretically uncorrelated, but individually auto-correlated:

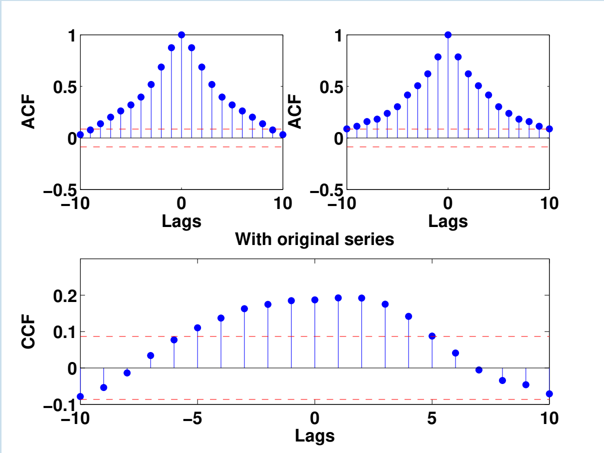
$$y[k] = \frac{1}{1 - 1.1q^{-1} + 0.28q^{-1}} e_1[k] \ \ \text{(AR(2))}; \quad u[k] = \frac{1}{1 - 0.8q^{-1}} e_2[k] \ \ \text{(AR(1))}$$

$$e_1[k] \sim \mathcal{N}(0, 1); \ \ e_2[k] \sim \mathcal{N}(0, 1); \quad \sigma_{eq_1 e_2}[l] = 0, \ \forall l$$

We compute $\rho_{yu}[l]$ from $N = 512$ observations as shown in Figure **??**.

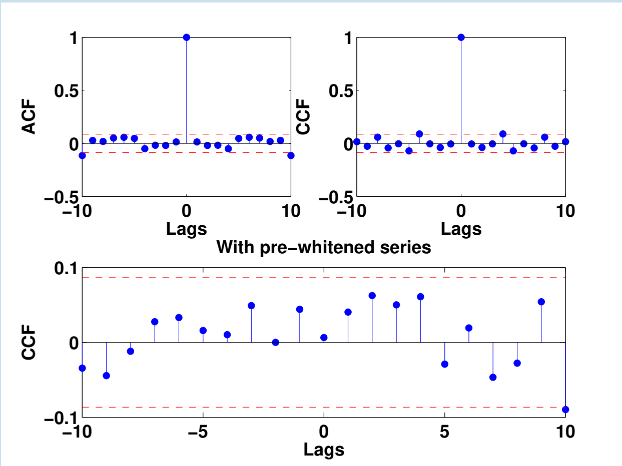
Example: . . . contd.

Although the series are truly uncorrelated, blindly applying the significance levels without considering the auto-correlation structure of the individual series can lead to erroneous conclusions.



Example: . . . contd.

Pre-whitening both the series followed by a CCF testing reveals the correct underlying relationship between the two signals $y[k]$ and $u[k]$.



Estimation of Auto-correlation function

Estimating ACF is a special case of the CCF. Therefore, the estimator and its properties simply inherit from that for the cross-correlation function.

Estimator:

$$\boxed{\hat{\rho}_{yy}[l] = \frac{\hat{\sigma}_{yy}[l]}{\hat{\sigma}_{yy}[0]}}; \quad \hat{\sigma}_{yy}[l] = \frac{1}{N} \sum_{k=l}^{N-l-1} (y[k] - \bar{y})(y[k-l] - \bar{y}), \qquad l \geq 0 \tag{30}$$

As in the case of CCF, the maximum lag up to which the ACF can be computed is $|l_{\text{max}}| = N - 1$.

Estimation of ACF ... contd.

By virtue of the preceding result, we have that, for *a Gaussian white-noise signal*,

$$\boxed{\begin{aligned} \hat{\rho}_{yy}[l] &\sim \text{As}\mathcal{N}(0, 1/N) && (31) \\ \implies 100(1-\alpha)\% \text{ sig. level for } \hat{\rho}_{yy}[l] &= \pm \frac{z_{\alpha/2}}{\sqrt{N}} && (32) \end{aligned}}$$

The above result forms the basis for **a test of whiteness** (test of predictability).

Tests of whiteness

Testing a given series for whiteness (zero temporal correlation) characteristics is an important step in TSA and identification. The hypothesis under examination is:

$$H_0 : \ \rho[l] = 0, \ \forall l \neq 0, \qquad H_a : \rho[l] \neq 0, \text{ for some non-zero lags}$$

Several methods are available for this purpose. For a detailed discussion, refer to Brockwell, 2002. We discuss two prominently used methods:

Tests of whiteness

1. **Sample auto-correlation test**: The setup for this test directly follows from (32). For a given series, compute the ACF estimates up to a maximum lag, say $l = 20$. Then, reject H_0 if at the ACF exceeds the $100(1-\alpha)\%$ significance level in (32). For instance, the 95% significance levels are $\pm 1.96\sqrt{N}$.
   A matter of concern in implementing this test is that the probability of $|\hat{\rho}[l]| > 1.96\sqrt{N}$ at a given lag depends on $\hat{\rho}[l]$ at other lags.

   See Box, Jenkins, and Reinsel, 2008 for a clear illustration of this test.

Portmanteau test

2. **Box-Ljung-Pierce test**: Or the *portmanteau test*. This method is superior to the above test because it *collectively* examines the ACF estimates over a range of lags. Under H_0, the sample ACF coefficients possess a Gaussian distribution; therefore the sum-squared estimates follow a χ^2 distribution.
   With this basic idea, the following statistic is constructed:

$$Q = N(N+2)\sum_{l=1}^{L} \frac{\hat{\rho}_{yy}^2[l]}{N-l} \tag{33}$$

Example: Whiteness test

Whiteness test of a series

Given $N = 512$ observations of a series, whose snapshot if displayed in Figure 1, we would like to test for its whiteness.

The sample ACF and the significance levels are shown in Figure 2. Based on the sample ACF test, the null hypothesis that the given series is white is rejected at $\alpha = 0.05$ significance level since the ACF is outside the 95% significance band at least at one non-zero lag.

Example: Whiteness test . . . contd.

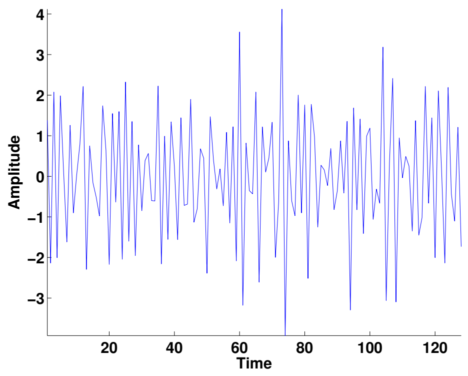


Figure 1: Snapshot of the series

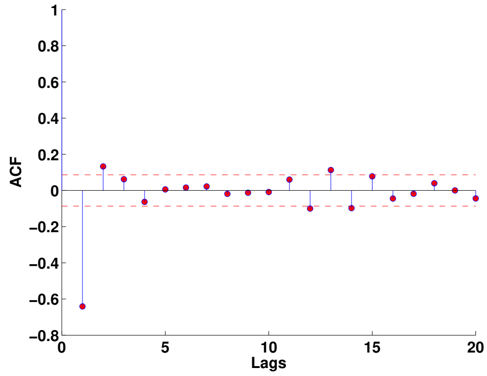


Figure 2: Sample ACF

Example: Whiteness test . . . contd.

The BLP Q statistic (with $L = 20$) for the given series is 250.6922, a value much higher than the critical value of 31.41 obtained from the $\chi^2_{20,\alpha/2}$ distribution where $\alpha = 0.05$. Thus, the null hypothesis that the series is white is rejected at $\alpha = 0.05$ significance level.

The ACF plot is in fact suggestive of a MA(2) model for the series. However, such guesses should be confirmed with a more rigorous analysis.

R commands

Listing 2: R commands for estimating cross- and auto-correlation

```

# Computing ACF and PACF
acf, pacf

# Computing CCF
ccf

# B-L Whiteness test
Box.test
```


Bibliography I

Box, G. E., G. M. Jenkins, and G. C. Reinsel (2008). *Time Series Analysis: Forecasting and Control*. New York, USA: John Wiley & Sons, Inc.

Brockwell, P. (2002). *Introduction to Time-Series and Forecasting*. New York, USA: Springer-Verlag.

Brockwell, P. and R. Davis (1991). *Time Series: Theory and Methods*. New York, USA: Springer.

Huber, P. J. and E. M. Ronchetti (2009). *Robust Statistics*. Probability and Statistics. New York, USA: John Wiley & Sons, Inc.

Pearson, R. K. (2011). *Exploring Data: in Engineering, the Sciences and Medicine*. Canada: Oxford University Press.

Pham-Gia, T. and T. Hung (2001). The mean and median absolute deviations. *Mathematical and Computer Modelling*, 24, pp. 921–936.

Bibliography II

Rousseeuw, P. J. and C. Croux (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, 88 (424), pp. 1273–1283.

Shumway, R. and D. Stoffer (2006). *Time Series Analysis and its Applications*. New York, USA: Springer-Verlag.

Tangirala, A. K. (2014). *Principles of System Identification: Theory and Practice*. Boca Raton, FL, USA: CRC Press, Taylor & Francis Group.